



# HOPMA: Boosting protein functional dynamics with colored contact maps

Elodie Laine, Sergei Grudinin

## ► To cite this version:

Elodie Laine, Sergei Grudinin. HOPMA: Boosting protein functional dynamics with colored contact maps. *Journal of Physical Chemistry B*, 2021, 125 (10), pp.2577-2588. 10.1021/acs.jpcb.0c11633 . hal-03091790

**HAL Id: hal-03091790**

**<https://inria.hal.science/hal-03091790>**

Submitted on 31 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial| 4.0 International License

# HOPMA: Boosting protein functional dynamics with colored contact maps

Elodie Laine<sup>a,1</sup> and Sergei Grudinin<sup>b,1</sup>

<sup>a</sup>Sorbonne Université, CNRS, IBPS, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), 75005 Paris, France; <sup>b</sup>Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000 Grenoble, France

In light of the recent very rapid progress in protein structure prediction, accessing the multitude of functional protein states is becoming more central than ever before. Indeed, proteins are flexible macromolecules, and they often perform their function by switching between different conformations. However, high-resolution experimental techniques such as X-ray crystallography and cryogenic electron microscopy can catch relatively few protein functional states. Many others are only accessible under physiological conditions in solution. Therefore, there is a pressing need to fill this gap with computational approaches.

We present HOPMA, a novel method to predict protein functional states and transitions using a modified elastic network model. The method exploits patterns in a protein contact map, taking its 3D structure as input, and excludes some disconnected patches from the elastic network. Combined with nonlinear normal mode analysis, this strategy boosts the protein conformational space exploration, especially when the input structure is highly constrained, as we demonstrate on a set of more than 400 transitions. Our results let us envision the discovery of new functional conformations, which were unreachable previously, starting from the experimentally known protein structures.

The method is computationally efficient and available at <https://github.com/elolaine/HOPMA> and <https://team.inria.fr/nano-d/software/nolb-normal-modes>.

Elastic Network Model | Protein contact map | Conformational dynamics  
| Functional motion | NMA | Nonlinear normal mode | Structural transition

## 1. Introduction

Proteins and their complexes are intrinsically flexible, and this flexibility is tightly linked to their biological functions (1, 2). The shape a protein adopts in solution and the way it moves is governed by a myriad of interatomic forces within the protein and with the surrounding solvent. Despite this high complexity, many motions relevant to protein functioning can be fairly approximated by a few low-frequency modes computed using the Normal Mode Analysis (NMA) and characteristic of the protein's geometrical shape (3–7). This observation has led to the development of many NMA-based methods to predict crystallographic temperature factors (8–12), refine crystallographic structures (13, 14), interpret or fit atomistic structures into cryo-EM maps (15–24) or one-dimensional scattering profiles (25–27), and predict the structures of protein complexes in combination with docking (1, 28–38), among other applications.

While the NMA formalism may be used with standard semi-empirical potential functions, highly simplified potentials, called *Elastic Network Models*, where neighboring atoms are connected by harmonic springs yield a good approximation of the low-frequency normal modes and are thus commonly used

(3–5). They have proven useful in modeling protein transitions (39–45), sampling protein conformational space (34, 46–49), partitioning proteins into rigid bodies (50–52), scanning single and double mutations (53–57), modeling allosteric signal propagation (58–60), and more. They have also helped to characterize the relationship between protein structural dynamics and the evolution of protein sequences (61–66).

The suitability of the NMA to model conformational dynamics crucially depends on the type of motions involved and on the protein 3D shape used to compute the modes (67). Typically, less constrained conformations encode more information about the dynamical potential of the protein than more constrained conformations. This is particularly visible when predicting a conformational transition from a closed to an open state. Indeed, opening the protein is significantly more difficult than closing it (3, 68). Lowering the distance cutoff used to determine whether two atoms are connected in the network may help to partially solve the problem. However, there is no universal optimum value applicable to any protein, and even within a given protein, it may be far from trivial to set a value that alleviates the constraints in some regions while avoiding sub-critical connectivity in others.

Several works have proposed to modulate the spring's stiffness depending on the type of interatomic connections or the local packing density (10, 69, 70). The resulting hybrid or multi-scale networks yield more accurate predictions of the temperature factors, but this gain is at the expense of a loss of universality and simplicity. Complications may arise, for example, from asymmetries in the definition of the network (70). Another strategy consists in updating the elastic network iteratively, either by re-building a network from a conformation obtained by deforming the initial network (68, 71), or by modifying the equilibrium distances on-the-fly, for example to fit the network into a low-resolution map (18, 72).

Here, we report on a method enhancing protein elastic networks dynamical potential by exploiting patterns in protein contact maps. We build smoothed binarized colored contact maps from inter-residue distances and identify the patches in these maps that correspond to small contact areas between protein segments. We remove these patches and compute negative images of the filtered contact maps defining pairs of protein regions that should not be linked by any spring. Contrary to protein domain detection methods (73–76), we do not attempt to partition the protein in a set of rigid bodies but focus on detecting the interatomic connections that lock the protein and hinder its motions. Our approach, called HOPMA (proteiN cOloRed MAPs), is conceptually simple,

<sup>1</sup> E.L. (Elodie Laine) and S.G. (Sergei Grudinin) contributed equally to this work. E-mail: [elodie.laine@sorbonne-universite.fr](mailto:elodie.laine@sorbonne-universite.fr), [sergei.grudinin@inria.fr](mailto:sergei.grudinin@inria.fr)

has only a few parameters and relies solely on the 3D information of the input structure. We show that HOPMA facilitates protein conformational space sampling by computing more than 400 protein functional transitions. We systematically assess HOPMA's contribution to the description of the transitions by comparing the results obtained when starting from a modified protein elastic network, where the contacts identified by HOPMA have been removed, with those obtained from "classical" elastic networks, only defined based on a distance cutoff.

There are multiple ways to compute protein transitions with NMA, for example, using normal modes in internal coordinates (77–81). Here, we used the NOnLinear rigid Block (NOLB) NMA, a Cartesian nonlinear approach that extrapolates the instantaneous eigenmotions as a series of twists (82). NOLB is computationally very efficient, as it can be applied to systems as large as viral capsids or ribosomes at all-atom resolution on a personal laptop (82). It produces conformations with high stereochemical quality and is able to predict motions with different degrees of collectivity, from localized and disruptive motions to highly collectives ones (68). We place ourself in the context where we know the target structure and we use this knowledge to determine the amplitudes of the initial structure's normal modes. This set up allows obtaining the optimal (or close-to-optimal) transitions within our framework. In practical applications, when the target is not know, the sampling may be guided by additional information (83, 84), such as experimental small-angle scattering profiles (25–27, 85, 86), low-resolution electron density maps (15–22), intensities from crystallographic scattering experiments (87, 88), complementarity of the molecular shapes upon binding (28–36), or other type of data such as contacts inferred from coevolutionary signals extracted from protein sequences (89–91).

## 2. Materials and Methods

**Datasets.** We considered four reference protein benchmark sets, namely the iMod benchmark (92), the HR-PDNA187 benchmark (93), the Protein-Protein Docking Benchmark v5 (PPDBv5) (94), and the dataset reported in (author?) (90) (Table 1). We used the iMod benchmark as the validation set and the other benchmarks as testing sets.

**Opening/closing set.** The iMod benchmark, available at <http://chaconlab.org/multiscale-simulations/imod/imod-download/item/imod-benchmark>, was previously used to assess coarse-grained elastic network model-based flexible fitting methods (95). It comprises 23 proteins, each given in "open" and "closed" conformations, and represents a wide variety of macromolecular motions, predominantly hinge motions but also shear and other complex motions. The structures come from the molecular motions database MolMovDB (96). All of them have less than 3% Ramachandran outliers (as computed by the MolProbity program (97)), do not have any broken chain or missing atom.

**Unbound-to-bound sets.** The HR-PDNA187 benchmark, available at <http://www.lcqb.upmc.fr/PDNAbenchmarks/>, was recently compiled to assess methods predicting DNA-binding patches at the surface of proteins. It comprises 187 high-quality structures (better than 2.5 Å resolution) of protein-DNA complexes and covers all major groups of protein-DNA interactions. The proteins have less than 25% sequence identity. 82 complexes

come along with the unbound forms of the proteins. We extracted 11 monomeric unbound-bound (free-complexed) pairs with  $C_\alpha$  RMS deviations between the two states above 2 Å.

The PPDBv5 benchmark, available at <https://zlab.umassmed.edu/benchmark/>, is well suited for assessing the range of applicability of flexible docking methods (98). It contains 230 protein complexes with at least one of the partners solved in both bound (complexed) and unbound (free) states. All structures have a resolution better than 3.25 Å, and some of them contain more than one chain. We extracted 68 single-chain proteins with  $C_\alpha$  root-mean-square deviation (RMSD) between the two states above 2 Å.

**Large diverse set.** The dataset reported in (author?) (90) was originally compiled to assess the ability of coevolutionary signals to guide the discovery of alternative functional protein conformations. It comprises 92 proteins sharing less than 70% sequence identity and corresponding to 224 protein chain structures. The motions exhibited by these structures are large (at least 3 Å RMSD) and of different types, namely open-close, rotation-close, rotation, concerted and other. The structures are grouped in 143 unique pairs, among which three are also found in the iMod benchmark (1akeA-4akeA, 1gggA-1wdnA, 1lstA-2laoA). For each pair, we considered both the forward and backward transitions, leading to a total of 286 transitions. We excluded four transitions because their initial structures (2ceaA, 3g7sB, 3rrmA and 1qmpB) displayed missing residues in the region serving as a hinge for the motion. Hence, the two parts of the protein moving with respect to each other were not connected by the protein backbone. We refer to this set as AlterBench.

**The protein cOLOred MAPs (HOPMA) algorithm.** HOPMA takes as input a protein 3D structure and gives as output a set of protein region pairs defining links that should be excluded in the elastic network model representing the input structure.

**From distances to contact propensities.** Given a protein 3D structure, we first compute all pairwise  $C_\alpha$ - $C_\alpha$  distances between the protein residues. Then, we convert these distances into contact propensities ranging between 0 and 1. The contact propensity  $p_{ij}$  between residues  $i$  and  $j$  is expressed as,

$$p_{ij} = \frac{d_{max} - \min(\max(d_{min}, d_{ij}), d_{max})}{d_{max} - d_{min}}, \quad [1]$$

where  $d_{ij}$  is the distance between the  $C_\alpha$  atoms of residues  $i$  and  $j$ , and  $d_{min}$  and  $d_{max}$  are predefined minimal and maximal distance bounds. Intuitively, a pair of residue distant by less than  $d_{min}$  will be considered as very close to each other and their contact propensity will be equal to 1. By contrast, a pair of residues distant by more than  $d_{max}$  will be considered as very far from each other and their contact propensity will be null. Different cutoff distances, typically in the interval of 6–12 Å, have been proposed in the literature to define  $C_\alpha$ - $C_\alpha$  contacts. Here, we set  $d_{min} = 3$  Å and  $d_{max} = 15$  Å.

**Detecting contiguous interaction patches.** Next, we proceed through a smoothing of the contact propensity matrix. Specifically, the smoothed value  $\tilde{p}_{ij}$  for the residue pair  $(i, j)$  is computed as the averaged contact propensity between the

**Table 1. Benchmark sets**

	#(structures)	#(transitions)	RMSD range (in Å)	types of transitions
iMod (92)	23	46	1.28 - 13.55	opening/closing
HR-PDNA187 (93)	11	11	2.22 - 28.37	DNA binding
PPDBv5 (94)	68	68	2.01 - 32.79	protein binding
AlterBench (90)	224	282	3.85 - 34.49	diverse

residues and their neighbours in the protein sequence:

$$\tilde{p}_{ij} = \frac{1}{(k+1)^2} \sum_{\substack{l \in [i-k, i+k] \\ m \in [j-k, j+k]}} p_{lm}. \quad [2]$$

We empirically set the size of the smoothing window to  $k = 2$ , such that the average contact propensities are computed over 25 cells of the contact propensity matrix  $P$ . In practice, as the matrix  $P$  is symmetrical, the smoothed values are computed only for the cells located in its lower triangle.

Finally, we convert the smoothed real-value matrix  $\tilde{P}$  into a binarised matrix  $B$ , depending on a cutoff value  $p_{cut}$ ,

$$b_{ij} = \begin{cases} 1, & \text{if } \tilde{p}_{ij} \geq p_{cut}, \\ 0, & \text{otherwise.} \end{cases} \quad [3]$$

In this binary matrix, the cells with a value of 1 are considered as active and the others as inactive. We further colour the matrix by detecting contiguous patches of active cells and labelling each active cell with a unique patch identifier using a fast-scanning version of the Hoshen–Kopelman algorithm, also called the connected component labelling 2-pass algorithm (99, 100).

**Defining the set of excluded contacts.** For each input structure, we compute two coloured maps with two different  $p_{cut}$  values, namely 0.001 and 0.1. The first value will lead to activating an  $(i, j)$  cell (*i.e.* setting its value to 1) whenever at least one of the residue pairs in the associated 25-cell window  $w = \{(l, k) \mid i - k \leq l \leq i + k, j - k \leq m \leq j + k\}$  is distant by less than 15 Å. The second value defines a more stringent activation criterion, which will be met for example when 3 residue pairs in the window are very close to each other (C $\alpha$  atoms distant by less than 5 Å). Within each coloured map, we select the contiguous patches smaller than  $(k+1)^2$  cells, where  $k$  is the size of the smoothing window. Here, as  $k = 2$ , we select patches up to 625 cells, which typically corresponds to an interaction between two supersecondary structures. We filter out the selected patches from the binary contact matrix  $B$  by inactivating the corresponding cells (*i.e.* setting their value to zero). We compute the final matrix as the average of the two filtered matrices  $B_{0.1}^{filt}$  and  $B_{0.001}^{filt}$  and we extract the set  $\mathcal{S} = \{(s_1, s_2), (s_3, s_4), \dots\}$  of protein segment pairs defining the inactive regions (values lower than 1) of that matrix.

**Computing non-linear-normal modes-driven protein transitions.** We predict protein transitions using the NOLB method (68, 82). Starting from a protein 3D structures, NOLB builds an elastic network model where any two atoms closeby in 3D space are linked by a harmonic spring. The model is anisotropic and comprises all the atoms of the protein (101, 102). It has

the following potential function,

$$V(q) = \sum_{d_{ij}^0 < R_c} \frac{\gamma}{2} (d_{ij} - d_{ij}^0)^2, \quad [4]$$

where  $d_{ij}$  is the distance between the  $i^{th}$  and the  $j^{th}$  atoms,  $d_{ij}^0$  is the reference distance between these atoms, as found in the original structure,  $\gamma$  is the spring constant, and  $R_c$  is a cutoff distance, set at 5 Å by default. Optionally, NOLB can modify the network by removing the links formed between some pairs of protein regions given as input. We use this functionality to account for the results generated by HOPMA in the construction of the elastic network.

NOLB then diagonalizes the rotation translation blocks (RTB)-projected mass-weighted Hessian matrix of this potential function to compute a set of eigenvectors or normal modes. The RTB technique is used to reduce the dimensionality of the diagonalization problem, by considering individual or several consecutive amino residues as rigid blocks (103–105). NOLB then deforms the input structure by extrapolating the instantaneous linear velocities and instantaneous angular velocities composing the eigenvectors to a series of twists (82). This guarantees a high stereo-chemical quality of the predicted conformations. To predict a transition from the *input* structure to some *target* structure, NOLB computes the displacement between the two structures and determines the optimal normal mode amplitudes (68). The transition can be optionally decomposed into smaller steps, each corresponding to a total linear deformation RMSD of 0.1 Å. The algorithm stops when the maximum number of iterations is exceeded (100 by default), or if the relative deformation becomes smaller than a tolerance of  $1e-6$ . NOLB also implements an iterative scheme where, at each iteration, a new elastic network model is built, representing the last predicted conformation, and the corresponding normal modes are computed (68).

**Assessing the predicted transitions.** To assess the quality of the computed transitions, we measure the extent to which they cover the conformational deviation between the aligned starting and target states. Transition coverage is computed as

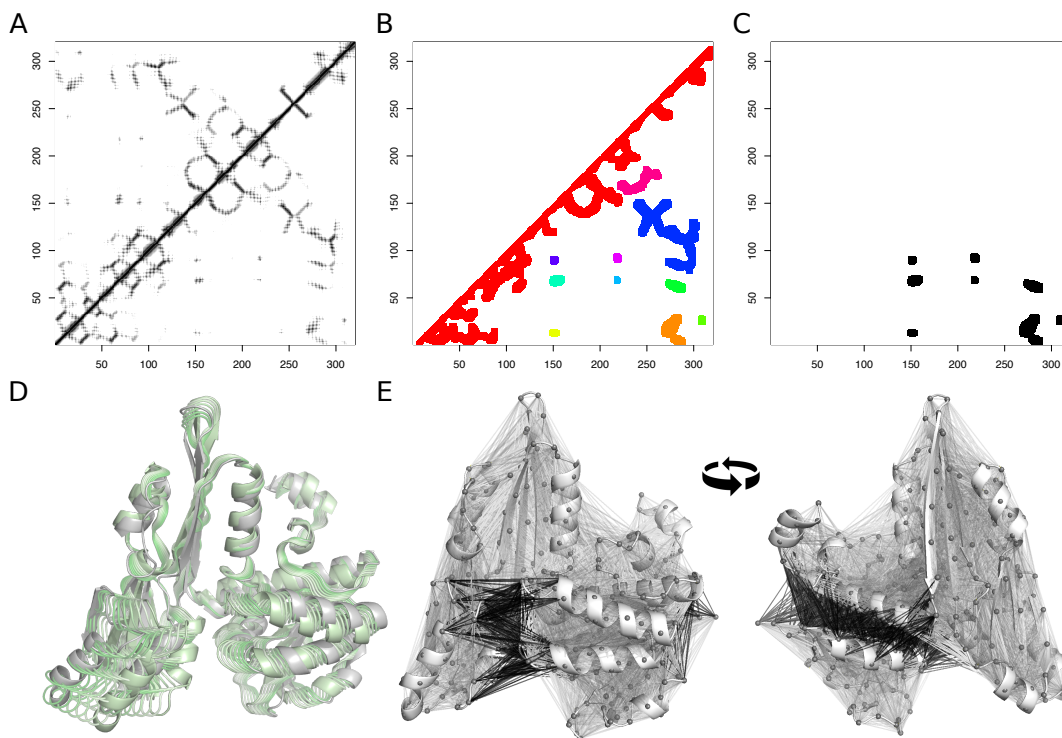
$$\text{Coverage} = \frac{\text{RMSD}_i - \text{RMSD}_f}{\text{RMSD}_i}, \quad [5]$$

where  $\text{RMSD}_i$  is the initial root mean square deviation between the starting and target structures, and  $\text{RMSD}_f$  is the deviation between the final structure obtained from the computed transition and the target structure. The coverage varies between 0 (null prediction) and 1 (perfect prediction).

### 3. Results and Discussion

**A colored-map-based approach to alleviate protein structural constraints.** We model a protein 3D structure as a graph  $\mathcal{G} =$





**Fig. 1. Principle of the method.** The closed conformation of the diaminopimelic acid dehydrogenase (PDB code: 1dapB) is taken as an illustrative example. **A.** Distance matrix computed between  $C\alpha$  atoms. The darker the dot, the closer the residues. **B.** Coloured binarised contact propensity matrix. Each colour indicates a contiguous patch of active cells in the matrix. **C.** Selection of the eight smallest patches. **D.** Nonlinear-normal-mode-driven transition toward the open conformation of the enzyme (PDB code: 3dapA). The initial and intermediate conformations are in green while the target conformation is in grey. The contacts corresponding to the eight identified patches were removed from the initial structure prior to the calculation of the transition. **E.** Mapping of the contact matrix shown in panel B onto the 3D structure of the protein. The contacts corresponding to the eight identified patches are highlighted in black, while the other ones are in grey.

$(\mathcal{V}, \mathcal{E})$ , where each node  $v \in \mathcal{V}$  corresponds to a protein atom and each edge in  $\mathcal{E}$  connecting  $v$  to  $v'$  represents an harmonic spring linking the corresponding atoms. The edges' weights indicate the spring constants. The simplest way to build such a graph, called a protein elastic network, is to determine the set of edges  $\mathcal{E}$  based on the interatomic pairwise euclidian distances and to set a unique value for the spring constant. Specifically, two nodes  $v$  and  $v' \in \mathcal{V}$  will be linked by a fixed-weight edge if they are distant by less than a certain distance cutoff, typically comprised between 3.5 Å and 15 Å. Different protein shapes will lead to different network topologies encoding different information about the dynamical potential of the protein.

Here, we propose to build an elastic network that only partially matches the protein structure to maximise its dynamical potential. To decide whether two nodes  $v$  and  $v'$  should be linked by an edge, we account for the 3D proximity of the corresponding atoms and also for their environment (Fig. 1A-C). More precisely, any two connected atoms in the network are closeby in 3D space (distant by less than 5 Å) and located in an extended contact area between two contiguous protein segments. We detect contact areas between protein segments as contiguous patches in a smoothed binarized *contact propensity map* (Fig. 1B) computed from the  $C\alpha$ - $C\alpha$  interatomic distances (Fig. 1A). These patches can be seen as connected components in a graph encoding 3D proximity at the residue level. In a globular protein, the largest patch is located around the diagonal and corresponds to the contiguous area defined by local contacts along the backbone (Fig. 1B,

in red). The off-diagonal patches correspond to interactions between disjoint segments remote in the protein sequence (Fig. 1B, in other colors). We allow building edges within large patches, representing extended contact areas, and forbid the creation of edges between atom pairs located in small patches or outside of the patches (Fig. 1C-D).

We implemented our approach as an automated tool called HOPMA – protein cOLOred MAPs. We assessed the ability of HOPMA to reveal protein 3D structures dynamical potential on 402 protein functional transitions covering a large spectrum of motions.

**HOPMA facilitates protein structure opening.** As a proof-of-concept, we first applied HOPMA to a set of 23 proteins exhibiting opening-closing motions (Fig. 2). We computed the transitions between the closed and open forms, in forward and backward directions, using the very rapid NOnLinear rigid Block (NOLB) NMA. Given a initial-target structure pair, NOLB computed the normal modes of the initial structure and deformed it along a combination of the ten lowest-frequency modes (see *Materials and Methods*). We evaluated the quality of a predicted transition by computing its transition coverage as the percentage of RMS deviation between the initial and target structures covered by the prediction.

Figure 2 shows a double comparison, the first one between the closed-to-open (on the left) and open-to-closed (on the right) directions, the second one between the "classical" elastic network (in purple), where the links are defined using a distance cutoff (5 Å), and the modified network (in green),



**Fig. 2. Transition coverage for opening/closing motions.** Comparison of the coverages achieved by the NOLB nonlinear modes for 23 transitions (see *Materials and Methods*, iMod benchmark) between a closed state and an open state, in the forward direction (closed-to-open, on the left) and backward direction (open-to-closed, on the right). On top, all the atom pairs distant by less than 5 Å are linked by a spring. At the bottom, we used HOPMA to remove some links. The darker the color, the higher the RMSD between the open and closed structures.

where some links are removed according to HOPMA results. Accounting for HOPMA results in the definition of the elastic network (**Fig. 2**, in green) allows achieving a transition coverage in the closed-to-open direction almost equivalent to that of the open-to-closed direction (average  $\Delta Coverage = 5 \pm 10\%$ ). By contrast, the coverage difference between the two directions is of 16% on average when starting from the "classical" elastic networks (**Fig. 2**, in purple). The HOPMA-enhanced networks are substantially easier to open, leading to an average coverage of the transition to the open form of  $65 \pm 19\%$ , compared to  $53 \pm 15\%$  with the "classical" elastic networks (**Fig. 2**, *Closed-to-Open*). The highest improvement is observed for the diaminopimelic acid dehydrogenase (1dap-3dap) for which the coverage increases from 35% to 75% (**Fig. 1**). Moreover, HOPMA leads to a significant coverage improvement (by more than 5%) in the open-to-closed direction for three proteins (**Fig. 2**, *Open-to-Closed*), namely the maltodextrin-binding protein (1anf-1omp), the dna topoisomerase III (1d6m-1i7d), and the lactoferrin (1lfh-1lfg).

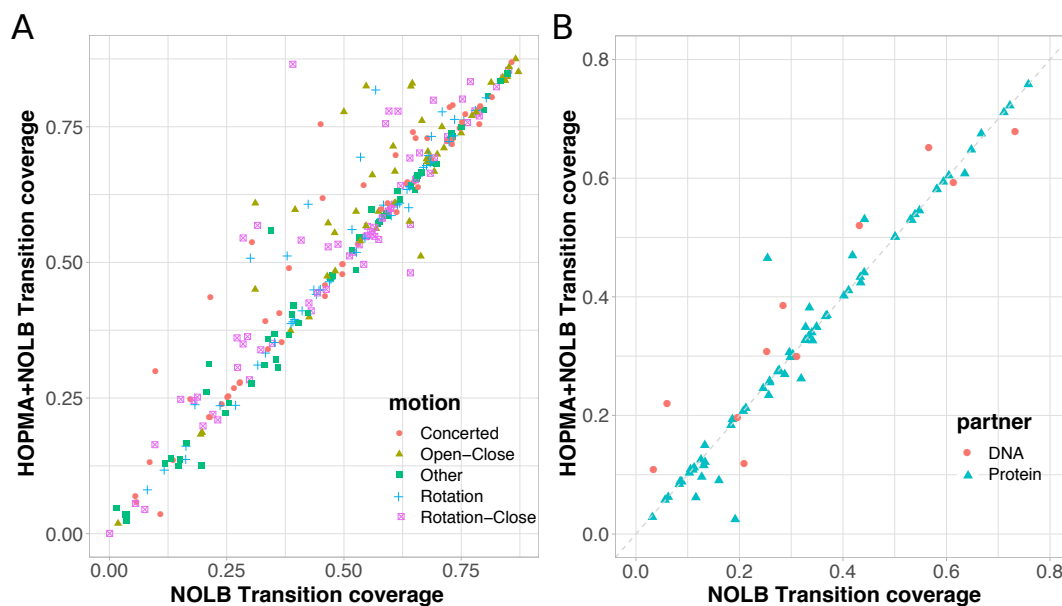
### HOPMA improves the prediction of a wide range of motions.

We then assessed HOPMA on a large and diverse testing set (AlterBench) comprising 143 protein structure pairs with RMS deviations ranging from 3 to 35 Å (**Fig. 3A**). All transitions were computed in the forward and backward directions, except for four cases, leading to a total of 282 transitions (see *Materials and Methods*). For the vast majority (98%) of cases, the HOPMA-enhanced networks lead to equivalent or better results compared to the "classical" networks (**Fig. 3A**). Moreover, HOPMA helps to describe a wide range of motions, including opening/closing motions, motions involving domain

rotations, concerted motions, and some more complex motions.

For example, HOPMA substantially improves the prediction of a rotation-close motion occurring in the pilus retraction motor PilT (2gsz), in both the forward and backward directions (**Fig. 4**). This hexameric ATPase is required for bacterial type IV pilus retraction and surface motility. In the hexamer, the ATPase core of one subunit extensively interacts with the N-terminal domain of the next subunit. As a result, the inter-domain distance and orientation vary from one subunit to another. The modelled transition takes place between two subunits (chains B and C) and implies some rearrangements of the contacts between the two domains (**Fig. 4**, compare the two colored maps). HOPMA identifies and removes the contacts between the two domains in both structures, leading to an increased transition coverage, from 39% to 86% in the forward direction (**Fig. 4**, on top), and from 62% to 78% in the backward direction (**Fig. 4**, at the bottom).

We also tested HOPMA on two other datasets, namely PPDBv5 and HR-PDNA187, dedicated to the study of protein-protein and protein-DNA interactions respectively (**Fig. 3B**). We selected 79 unbound-bound structure pairs with at least 2 Å RMS deviation between them, and we only focused on the prediction of the unbound-to-bound transitions (see *Materials and Methods*). As observed on the AlterBench set, the HOPMA-enhanced elastic networks lead to equivalent or better transition coverages compared to the "classical" elastic networks (**Fig. 3B**). However, the transitions are overall more difficult to predict compared to the AlterBench set (**Figure S1**). HOPMA significantly improved the predictions of nine functional transitions, three undergone upon binding to a protein partner and six upon binding to DNA (**Fig. 3B**). For



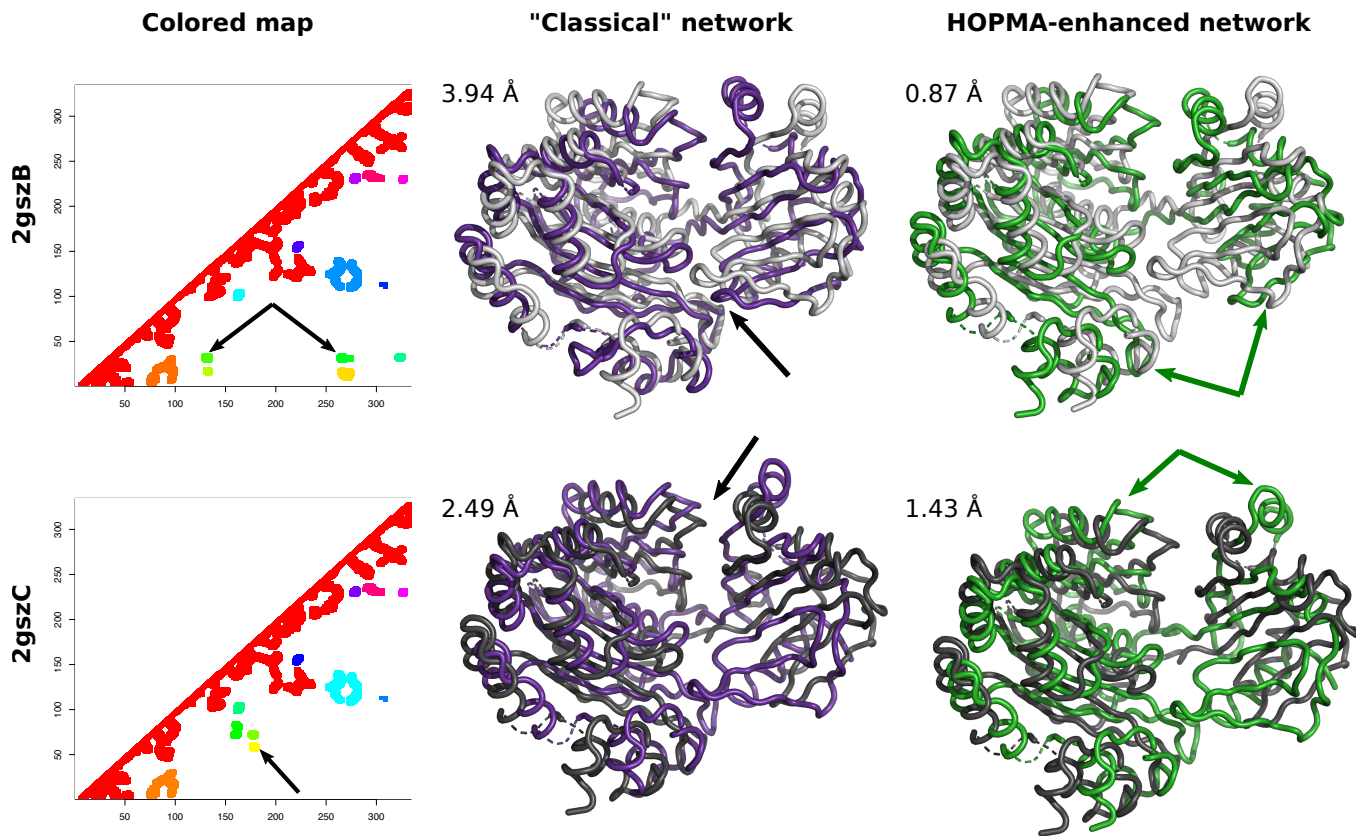
**Fig. 3. Large scale assessment of HOPMA.** The transition coverages achieved with the elastic networks defined using HOPMA are compared with those obtained with the "classical" networks (distance cutoff only). **A.** 282 transitions displaying a wide range of motions from the AlterBench dataset (see *Materials and Methods*). For each pair of structures defined in the set, we computed both the forward and backward transitions. **B.** 79 unbound-to-bound transitions from the PPDBv5 (protein-protein interactions) and HR-PDNA187 (protein-DNA interactions).

example, it increases the coverage of the transition of the human vinculin head upon binding to the vinculin tail from 25% to 47%. The conformational change between the initial and target structures concern 20% of the protein and amounts to 4.25 Å. The low coverage obtained with the "classical" elastic network can be explained by the presence of contacts with C-terminal extremity the locks the position of a helix-turn-helix. In the HOPMA-enhanced network, the contacts are removed and the helices are free to rotate (Fig. S2).

**HOPMA reduces the number of iterations required for very large DNA-induced conformational changes.** To describe very large transitions or transitions involving complex motions, it may be useful to re-iterate the NOLB calculation one or several times. At each iteration, a new elastic network model is computed from the last predicted conformation and a new transition is computed using its first ten normal modes (see *Materials and Methods*). We investigated whether HOPMA could also be useful in this multi-iterative scheme. We focused on the 11 transitions associated with DNA binding from the HR-PDNA187 dataset. We found that accounting for HOPMA results to define the elastic networks leads to conformations that are closer to the target, after three iterations, for the three very large transitions of the set (Fig. S3, initial RMSD > 10 Å). The highest gain is obtained for the Y-family DNA polymerase Dpo4 (2rdiA-1jxoA). This enzyme plays a crucial role in translesion synthesis and undergoes a dramatic conformational change upon DNA binding (Fig. 5). The transition implies a ~130 degrees rotation of the little finger domain relative to the polymerase core resulting in a RMS deviation of 11.47 Å between the unbound and bound structures. HOPMA marked the contacts between the two domains as to be removed and this resulted in a transition coverage of 38% instead of 28% with the "classical" network after the first iteration. Re-iterating the transition computa-

tion twice produces a final conformation 1.80 Å away from the target that can accommodate DNA without backbone clashes (Fig. 5, in purple). By contrast, the same procedure without using HOPMA produces a more distant conformation, with visible distortions in the little finger domain, that cannot accommodate DNA (Fig. 5).

**HOPMA can be useful to explore a protein conformational space.** Beyond the prediction of a transition between two functional states, we investigated whether HOPMA could help to navigate within an ensemble of protein conformations. We took the multi-functional calcium sensor calmodulin as a case study. Calmodulin structural flexibility has been extensively characterized and shown to play a key role in its ability to bind a tremendous amount of proteins and peptides (106). We computed all the pairwise transitions between 27 conformations of the protein, both in the forward and in the backward directions (Fig. 6). The HOPMA-enhanced networks systematically led to equivalent or better predictions compared to the "classical" networks (Fig. 6B). One can clearly identify seven conformers (numbered 1, 5, 12, 18, 20, 25 and 27) whose transitions to the other conformers are significantly facilitated. In these conformers, the two domains of the protein establish substantial contacts that prevent them from changing their position and orientation with respect to each other. HOPMA detects the contacts and marks them as to be removed, enhancing the dynamical potential of the protein network. The most striking coverage gain is observed for the conformer 27, which corresponds to an X-ray crystallographic structure of the protein complexed with a substrate peptide and four calcium ions (Fig. 6C-D). The conformation is very compact, with three contact areas between the two protein domains (Fig. 6C). While NOLB is unable to predict the transitions to the other conformers when starting with the "classical" networks (average Coverage = 0.22), it covers 50%



**Fig. 4. Conformational change between two subunits of the pilus retraction motor PilT.** The experimental structures of the two subunits are shown as light grey (2gszB, on top) and dark grey (2gszC, at the bottom) cartoons. They deviate by 6.47 Å. The corresponding coloured maps computed by HOPMA are displayed on the left. The final conformations of the predicted transitions are shown as purple ("classical" network) and green (HOPMA-enhanced network) cartoons and superimposed onto the initial structure. Their RMS deviations to the target structures are given. The black arrows indicate some contacting regions in the initial structures, which are maintained in the conformations predicted from the "classical" networks. The green arrows highlight some opening motions correctly predicted in these regions when starting from the HOPMA-enhanced networks.

of the transitions, on average, when starting from the less constrained HOPMA-enhanced network (Fig. 6D, see also Fig. S4). Nevertheless, this is still significantly less than what is achieved among the other conformers, which correspond to free forms of the protein (average *Coverage* = 0.73, see Fig. S4). This difference may be explained by more subtle and more difficult to predict conformational changes between the free and complexed states that affect the orientation of the helix pairs (EF-hand motifs) within each domain and the associated solvent exposure of some hydrophobic residues.

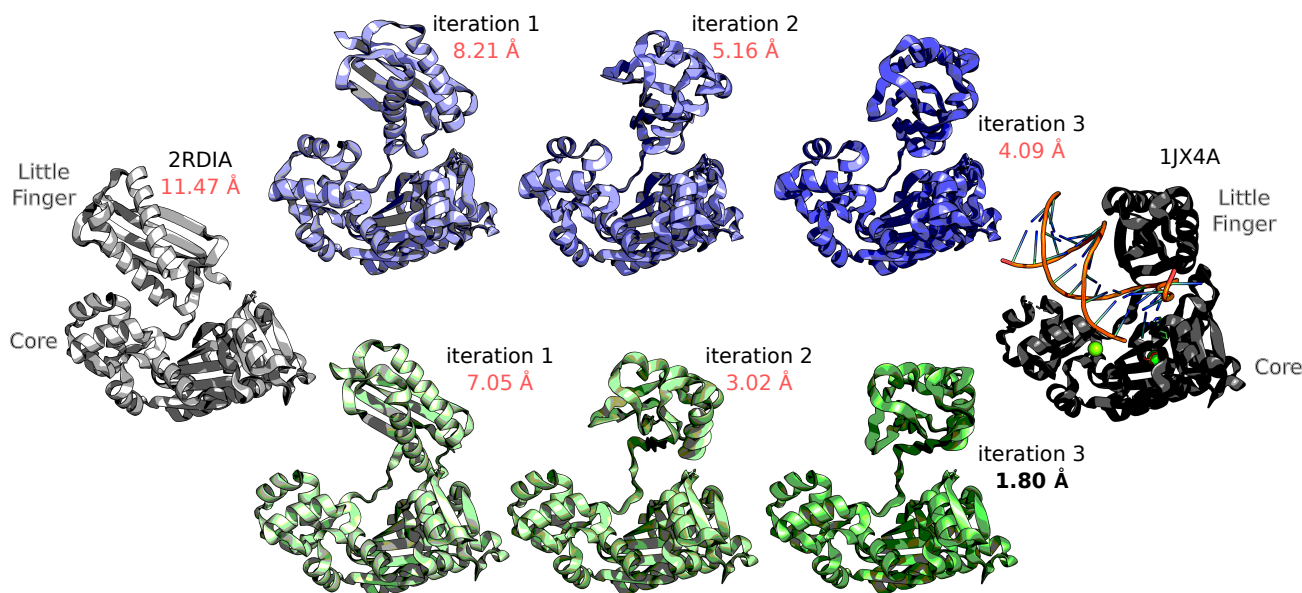
**Computational details and availability.** The HOPMA code was written in Python3 and R. All the calculations were performed on a 2,7 GHz Intel Core i7 CPU with 16 Go RAM. Given a protein structure, we used the following command to run HOPMA: "python3 \$HOPMA\_PATH/hopma.py prot -c chain", where *prot* is the name of the input PDB structure file without the extension, and *chain* is the name of the chain to analyze. Given two protein structures *protInitial* and *protTarget*, we used the following command to run NOLB: "NOLB protInitial protTarget -n 10 --nlin -m". To account for HOPMA results, we used the "--excl" option with the output of HOPMA. The computations are efficient, for example, predicting a transition for a protein of about 300 residues takes 12 seconds,

including 9 seconds to perform HOPMA analysis and 3 seconds to run NOLB. The execution time increases up to 81 seconds for a protein of about 1 000 residues (66 seconds for HOPMA and 15 seconds for NOLB). The efficiency of HOPMA can be improved by transferring it to C++. The method is available at <https://github.com/elolaine/HOPMA> and <https://team.inria.fr/nano-d/software/nolb-normal-modes>.

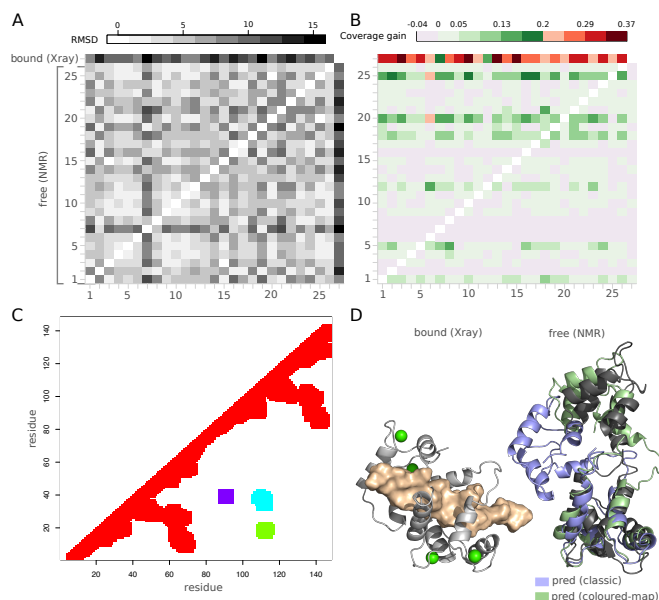
## 4. Conclusions

This work further develops our approach to predicting conformational changes in proteins and other macromolecules. Here, we implemented and carefully assessed the novel idea of modifying the initial elastic network by removing links corresponding to disconnected patches in the contact map. To detect these patches, we rely only on the 3D information encoded in the input protein structure. As expected, the modified networks led to significantly improved transitions, as we showed on a large set of examples taken from four diversified benchmarks. The method is user-friendly and computationally efficient. This work opens perspectives for sampling new protein functional conformations that were not accessible before, starting from experimentally determined structures.





**Fig. 5. DNA Polymerase conformational change upon binding to DNA and an incoming nucleotide.** The unbound structure (2rdia) is shown on the left in grey while the structure bound to DNA and a nucleotide (1jx4a) is shown on the right in black. The little finger and core domains are labeled. The conformations in between were predicted by NOLB, after 1, 2 and 3 iterations respectively. The conformations in purple were obtained from "classical" elastic networks, while those in green were obtained from less constrained networks, where the links identified by HOPMA were removed. The RMS deviations to the target are given and highlighted in red when the backbone of the conformation clashes with the DNA.



**Fig. 6. Exploration of Calmodulin conformational space.** **A.** Pairwise RMS deviations between 27 conformations of calmodulin, corresponding to 26 NMR models of the free protein (1cfc and 1cfd) and one crystallographic structure of the protein bound to a peptide and some calcium ions (2bbm). **B.** Transition coverage gain when predicting all pairwise transitions starting from HOPMA-enhanced elastic networks with respect to "classical" elastic networks. **C.** Colored map computed by HOPMA for the bound structure (2bbm). **D.** On the left, the bound structure is displayed as a light grey cartoon with the peptide in colored surface and the calcium ions as green spheres. On the right, the free structure (first monomer of 1cfc) is displayed as a dark grey cartoon, while the final conformations predicted from elastic networks representing the bound structure are displayed as colored cartoons. The conformation predicted from the "classical" network is shown in purple while the one generated from the HOPMA-enhanced network is in green.

## 5. Acknowledgments

The authors are thankful to Vladimir Sorokin for inspiration.

- Ozdemir ES, Nussinov R, Gursoy A, Keskin O (2019) Developments in integrative modeling with dynamical interfaces. *Current opinion in structural biology* 56:11–17.
- Gunasekaran K, Nussinov R (2007) How different are structurally flexible and rigid binding sites? sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *Journal of molecular biology* 365(1):257–273.
- Tama F, Sanejouand YH (2001) Conformational change of proteins arising from normal mode calculations. *Protein Engineering* 14(1):1–6.
- Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical Review Letters* 77(9):1905.
- Krebs WG, et al. (2002) Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins* 48(4):682–695.
- Hayward S, Go N (1995) Collective variable description of native protein dynamics. *Annual Review of Physical Chemistry* 46(1):223–250.
- Zheng W, Wen H (2017) A survey of coarse-grained methods for modeling protein conformational transitions. *Current Opinion in Structural Biology* 42:24–30.
- Kondrashov DA, Van Wynsbeghe AW, Bannen RM, Cui Q, Phillips Jr GN (2007) Protein structural variation in computational models and crystallographic data. *Structure* 15(2):169–177.
- Mendez R, Bastolla U (2010) Torsional network model: normal modes in torsion angle space better correlate with conformation changes in proteins. *Physical Review Letters* 104(22):228103.
- Fuglebakk E, Reuter N, Hinsen K (2013) Evaluation of protein elastic network models based on an analysis of collective motions. *Journal of chemical theory and computation* 9(12):5618–5628.
- Zhou L, Liu Q (2014) Aligning experimental and theoretical anisotropic b-factors: water models, normal-mode analysis methods, and metrics. *The Journal of Physical Chemistry B* 118(15):4069–4079.
- Na H, Hinsen K, Song G (2020) The amounts of thermal vibrations and static disorder in protein x-ray crystallographic b-factors. *Authorea Preprints*.
- Delarue M, Dumas P (2004) On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proceedings of the National Academy of Sciences* 101(18):6957–6962.
- Lindahl E, Azuara C, Koehl P, Delarue M (2006) NOMAD-Ref: Visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.* 34(2):W52–W56.
- Tama F, Miyashita O, III CLB (2004) Normal Mode Based Flexible Fitting of High-Resolution Structure into Low-Resolution Experimental Data from Cryo-EM. *Journal of Structural Biology* 147(3):315–326.
- Tama F, Miyashita O, Brooks III CL (2004) Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *Journal of Molecular Biology* 337(4):985–999.



17. Suhre K, Navaza J, Sanejouand YH (2006) NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallographica Section D: Biological Crystallography* 62(9):1098–1100.
18. Schröder GF, Brunger AT, Levitt M (2007) Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* 15(12):1630–1641.
19. Schröder GF, Levitt M, Brunger AT (2012) Super-resolution biomolecular crystallography with low-resolution data. *Nature* 464(7292):1218.
20. Tan RKZ, Devkota B, Harvey SC (2008) YUP. SCX: coaxing atomic models into medium resolution electron density maps. *Journal of Structural Biology* 163(2):163–174.
21. Zheng W (2011) Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. *Biophysical Journal* 100(2):478–488.
22. López-Blanco JR, Chacón P (2013) iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *Journal of Structural Biology* 184(2):261–270.
23. Vilas J, et al. (2018) Advances in image processing for single-particle analysis by electron cryomicroscopy and challenges ahead. *Current Opinion in Structural Biology* 52:127–145.
24. Harastani M, Sorzano COS, Jonić S (2020) Hybrid electron microscopy normal mode analysis with scipion. *Protein Science* 29(1):223–236.
25. Gorba C, Miyashita O, Tama F (2008) Normal-mode flexible fitting of high-resolution structure of biological molecules toward one-dimensional low-resolution data. *Biophysical Journal* 94(5):1589–1599.
26. Wen B, Peng J, Zuo X, Gong Q, Zhang Z (2014) Characterization of protein flexibility using small-angle x-ray scattering and amplified collective motion simulations. *Biophysical Journal* 107(4):956–964.
27. Cheng P, Peng J, Zhang Z (2017) Saxes-oriented ensemble refinement of flexible biomolecules. *Biophysical Journal* 112(7):1295–1301.
28. Maschiach E, Nussinov R, Haim W (2010) Fiberdock: Flexible induced-fit backbone refinement in molecular docking. *Proteins Struct. Funct. Bioinf.* 78(6):1503–1519.
29. Venkatraman V, Ritchie DW (2012) Flexible protein docking refinement using pose-dependent normal mode analysis. *Proteins Struct. Funct. Bioinf.* 80(9):2262–2274.
30. Cavasotto CN, Kovacs JA, Abagyan RA (2005) Representing receptor flexibility in ligand docking through relevant normal modes. *Journal of the American Chemical Society* 127(26):9632–9640.
31. Lindahl E, Delarue M (2005) Refinement of docked protein–ligand and protein–dna structures using low frequency normal mode amplitude optimization. *Nucleic Acids Res.* 33(14):4496–4506.
32. Mustard D, Ritchie DW (2005) Docking essential dynamics eigenstructures. *Proteins Struct. Funct. Bioinf.* 60(2):269–274.
33. May A, Zacharias M (2008) Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking. *Proteins Struct. Funct. Bioinf.* 70(3):794–809.
34. Neveu E, et al. (2018) RapidRMSD: Rapid determination of RMSDs corresponding to motions of flexible molecules. *Bioinformatics* 34(16):2757–2765.
35. Moal IH, Bates PA (2010) SwarmDock and the use of normal modes in protein–protein docking. *International Journal of Molecular Sciences* 11:3623–3648.
36. Fiorucci S, Zacharias M (2010) Binding site prediction and improved scoring during flexible protein–protein docking with ATTRACT. *Proteins Struct. Funct. Bioinf.* 78(15):3131–3139.
37. Philot EA, Perahia D, Braz ASK, de Souza Costa MG, Scott LPB (2013) Binding sites and hydrophobic pockets in human thioredoxin 1 determined by normal mode analysis. *Journal of structural biology* 184(2):293–300.
38. Moroy G, et al. (2015) Sampling of conformational ensemble for virtual screening using molecular dynamics simulations and normal mode analysis. *Future medicinal chemistry* 7(17):2317–2331.
39. Zheng W, Thirumalai D (2009) Coupling between normal modes drives protein conformational dynamics: illustrations using allosteric transitions in myosin ii. *Biophysical Journal* 96(6):2128–2137.
40. Seo S, Kim MK (2012) Kosmos: a universal morph server for nucleic acids, proteins and their complexes. *Nucleic acids research* 40(W1):W531–W536.
41. Krüger DM, Ahmed A, Gohlke H (2012) Nmsim web server: integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. *Nucleic acids research* 40(W1):W310–W316.
42. Das A, et al. (2014) Exploring the conformational transitions of biomolecular systems using a simple two-state anisotropic network model. *PLoS Comput Biol* 10(4):e1003521.
43. Orellana L, Yoluk O, Carrillo O, Orozco M, Lindahl E (2016) Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. *Nature communications* 7(1):1–14.
44. Orellana L, Gustavsson J, Bergh C, Yoluk O, Lindahl E (2019) eBDIMS server: protein transition pathways with ensemble analysis in 2D-motion spaces. *Bioinformatics* 35(18):3505–3507.
45. Wang A, Zhang D, Li Y, Zhang Z, Li G (2019) Large-scale biomolecular conformational transitions explored by a combined elastic network model and enhanced sampling molecular dynamics. *The Journal of Physical Chemistry Letters* 11(1):325–332.
46. Zheng W, Brooks BR (2005) Normal-modes-based prediction of protein conformational changes guided by distance constraints. *Biophysical Journal* 88(5):3109–3117.
47. Kurkuoglu Z, Bahar I, Doruker P (2016) Clustenn: Enm-based sampling of essential conformational space at full atomic resolution. *Journal of chemical theory and computation* 12(9):4549–4562.
48. Kurkuoglu Z, Doruker P (2016) Ligand docking to intermediate and close-to-bound conformers generated by an elastic network model based algorithm for highly flexible proteins. *PLoS one* 11(6):e0158063.
49. Kurkuoglu Z, Bonvin AM (2020) Pre- and post-docking sampling of conformational changes using clustenn and haddock for protein–protein and protein–dna systems. *Proteins: Structure, Function, and Bioinformatics* 88(2):292–306.
50. Emecli U, Schneidman-Duhovny D, Wolfson H, Nussinov R, Haliloglu T (2008) Hinge-Prot: Automated prediction of hinges in protein structures. *Proteins Struct. Funct. Bioinf.* 70(4):1219–1227.
51. Schneidman-Duhovny D, Nussinov R, Wolfson HJ (2007) Automatic prediction of protein interactions with large scale motion. *Proteins Struct. Funct. Bioinf.* 69(4):764–773.
52. Kundu S, Sorensen DC, Phillips Jr GN (2004) Automatic domain decomposition of proteins by a gaussian network model. *Proteins: Structure, Function, and Bioinformatics* 57(4):725–733.
53. Atilgan C, Gerek Z, Ozkan S, Atilgan A (2010) Manipulation of conformational change in proteins by single-residue perturbations. *Biophysical Journal* 99(3):933–943.
54. Rodrigues CH, Pires DE, Ascher DB (2018) Dynamut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic acids research* 46(W1):W350–W355.
55. Tiberti M, Pandini A, Fraternali F, Fornili A (2018) In silico identification of rescue sites by double force scanning. *Bioinformatics* 34(2):207–214.
56. Zhang PF, Su JG (2019) Identification of key sites controlling protein functional motions by using elastic network model combined with internal coordinates. *The Journal of chemical physics* 151(4):045101.
57. Echave J (2020) Fast and exact single and double mutation-response scanning of proteins. *bioRxiv*.
58. Wang J, et al. (2020) Mapping allosteric communications within individual proteins. *Nature communications* 11(1):1–13.
59. General IJ, et al. (2014) ATPase subdomain IA is a mediator of interdomain allostery in Hsp70 molecular chaperones. *PLoS Comput Biol* 10(5):e1003624.
60. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Des.* 2(3):173–81.
61. Echave J (2008) Evolutionary divergence of protein structure: the linearly forced elastic network model. *Chemical physics letters* 457(4–6):413–416.
62. Echave J (2012) Why are the low-energy protein normal modes evolutionarily conserved? *Pure and Applied Chemistry* 84(9):1931–1937.
63. Liu Y, Bahar I (2012) Sequence evolution correlates with structural dynamics. *Molecular biology and evolution* 29(9):2253–2263.
64. Bakan A, et al. (2014) Evol and prody for bridging protein sequence evolution and structural dynamics. *Bioinformatics* 30(18):2681–2683.
65. Tiwari SP, Reuter N (2018) Conservation of intrinsic dynamics in proteins—what have computational models taught us? *Current Opinion in Structural Biology* 50:75–81.
66. Liang Z, Verkhivker GM, Hu G (2020) Integration of network models and evolutionary analysis into high-throughput modeling of protein dynamics and allosteric regulation: theory, tools and applications. *Briefings in Bioinformatics* 21(3):815–835.
67. Ma J (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* 13(3):373–380.
68. Grudinin S, Laine E, Hoffmann A (2020) Predicting protein functional motions: an old recipe with a new twist. *Biophysical Journal*.
69. Ming D, Wall ME (2005) Allostery in a coarse-grained model of protein dynamics. *Physical Review Letters* 95(19):198103.
70. Xia K, Opron K, Wei GW (2015) Multiscale gaussian network model (mgnm) and multiscale anisotropic network model (mann). *The Journal of chemical physics* 143(20):11B616.1.
71. Murshudov GN, et al. (2011) Refmac5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* 67(4):355–367.
72. Schröder GF, Levitt M, Brunger AT (2014) Deformable elastic network refinement for low-resolution macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* 70(9):2241–2255.
73. Alexandrov N, Shindyalov I (2003) PDP: protein domain parser. *Bioinformatics* 19(3):429–430.
74. Zhou H, Xue B, Zhou Y (2007) DDOMAIN: Dividing structures into domains using a normalized domain–domain interaction profile. *Protein science* 16(5):947–955.
75. Gelly JC, de Brevin AG (2010) Protein Peeling 3D: new tools for analyzing protein structures. *Bioinformatics* 27(1):132–133.
76. Postic G, Ghouzam Y, Chebrek R, Gelly JC (2017) An ambiguity principle for assigning protein structural domains. *Science advances* 3(1):e1600552.
77. Wako H, Endo S (2013) Normal mode analysis based on an elastic network model for biomolecules in the protein data bank, which uses dihedral angles as independent variables. *Computational Biology and Chemistry* 44:22–30.
78. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P (2014) iMODS: internal coordinates normal mode analysis server. *Nucleic acids research* 42(W1):W271–W276.
79. Bastolla U, Dehouck Y (2019) Can conformational changes of proteins be represented in torsion angle space? a study with rescaled ridge regression. *Journal of Chemical Information and Modeling* 59(11):4929–4941.
80. Frezza E, Lavery R (2019) Internal coordinate normal mode analysis: A strategy to predict protein conformational transitions. *The Journal of Physical Chemistry B* 123(6):1294–1301.
81. Diaz NC, Frezza E, Martin J (2020) Using normal mode analysis on protein structural models. how far can we go on our predictions? *Proteins*.
82. Hoffmann A, Grudinin S (2017) NOLB: Nonlinear rigid block normal-mode analysis method. *Journal of Chemical Theory and Computation* 13(5):2123–2134.
83. Vashisth H, Skiniotis G, Brooks III CL (2014) Collective variable approaches for single molecule flexible fitting and enhanced sampling. *Chemical Reviews* 114(6):3353–3365.
84. Srivastava A, Tiwari SP, Miyashita O, Tama F (2020) Integrative/hybrid modeling approaches for studying biomolecules. *Journal of Molecular Biology* 432(9):2846–2860.
85. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (saxs) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics* 40(3):191–285.
86. Grudinin S, Garkavenko M, Kazennov A (2017) Pepsi-saxs: an adaptive method for rapid and accurate computation of small-angle x-ray scattering profiles. *Acta Crystallographica Section D: Structural Biology* 73(5):449–464.
87. Suhre K, Sanejouand YH (2004) On the potential of normal-mode analysis for solving diffi-

- cult molecular-replacement problems. *Acta Crystallographica Section D: Biological Crystallography* 60(4):796–799.
88. McCoy AJ, Nicholls RA, Schneider TR (2013) Sceds: protein fragments for molecular replacement in phaser. *Acta Crystallographica Section D: Biological Crystallography* 69(11):2216–2225.
  89. Morcos F, Jana B, Hwa T, Onuchic JN (2013) Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proceedings of the National Academy of Sciences* 110(51):20533–20538.
  90. Sfriso P, et al. (2016) Residues coevolution guides the systematic identification of alternative functional conformations in proteins. *Structure* 24(1):116–126.
  91. Feng J, Shukla D (2020) Fingerprintcontacts: Predicting alternative conformations of proteins from coevolution. *The Journal of Physical Chemistry B* 124(18):3605–3615.
  92. Lopéz-Blanco JR, Garzón JI, Chacón P (2011) iMod: Multipurpose normal mode analysis in internal coordinates. *Bioinformatics* 27(20):2843–2850.
  93. Corsi F, Lavery R, Laine E, Carbone A (2020) Multiple protein-dna interfaces unravelled by evolutionary information, physico-chemical and geometrical properties. *PLoS computational biology* 16(2):e1007624.
  94. Vreven T, et al. (2015) Updates to the integrated protein-protein interaction benchmarks: Docking benchmark version 5 and affinity benchmark version 2. *Journal of Molecular Biology* 427(19):3031–3041.
  95. Tekpinar M (2018) Flexible fitting to cryo-electron microscopy maps with coarse-grained elastic network models. *Molecular Simulation* pp. 1–9.
  96. Echols N, Milburn D, Gerstein M (2003) MolMovDB: Analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Research* 31(1):478–482.
  97. Chen VB, et al. (2010) Molprobity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Structural Biology* 66(1):12–21.
  98. Dobbins SE, Lesk VI, Sternberg MJE (2008) Insights into protein flexibility: the relationship between normal modes and conformational change upon protein–protein docking. *Proceedings of the National Academy of Sciences of the United States of America* 105(30):10390–10395.
  99. Hoshen J, Kopelman R (1976) Percolation and cluster distribution. i. cluster multiple labeling technique and critical concentration algorithm. *Physical Review B* 14(8):3438.
  100. He L, Chao Y, Suzuki K (2008) A run-based two-scan labeling algorithm. *IEEE transactions on image processing* 17(5):749–756.
  101. Doruker P, Atilgan AR, Bahar I (2000) Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to  $\alpha$ -amylase inhibitor. *Proteins: Structure, Function and Bioinformatics* 40(3):512–524.
  102. Atilgan AR, et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal* 80(1):505–515.
  103. Durand P, Trinquier G, Sanejouand YH (1994) A new approach for determining low-frequency normal modes in macromolecules. *Biopolymers* 34(6):759–771.
  104. Tama F, Gadea FX, Marques O, Sanejouand YH (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Structure, Function and Bioinformatics* 41(1):1–7.
  105. Li G, Cui Q (2002) A coarse-grained normal mode approach for macromolecules: An efficient implementation and application  $\text{Ca}^{2+}$ -ATPase. *Biophysical Journal* 83(5):2457–2474.
  106. Yamniuk AP, Vogel HJ (2004) Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. *Molecular biotechnology* 27(1):33–57.